Available online at www.sciencedirect.com
ScienceDirect
Journal of Multivariate Analysis 99 (2008) 815–833
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

The one- and multi-sample problem for functional data with application to projective shape analysis

A. Munk[a,1], R. Paige[b,*], J. Pang[b], V. Patrangenaru[c,2], F. Ruymgaart[b,3]

[a] *Institute for Mathematical Stochastics, Gottingen University, Germany*
[b] *Department of Mathematics and Statistics, Texas Tech University, USA*
[c] *Department of Statistics, Florida State University, USA*

Received 22 May 2006
Available online 21 April 2007

Abstract

In this paper tests are derived for testing neighborhood hypotheses for the one- and multi-sample problem for functional data. Our methodology is used to generalize testing in projective shape analysis, which has traditionally involving data consisting of finite number of points, to the functional case. The one-sample test is applied to the problem of scene identification, in the context of the projective shape of a planar curve.

AMS 2000 subject classification: primary 62G35; secondary 62H35

Keywords: Functional data; Precise hypotheses; Validation of hypotheses; Image analysis; Projective shape data analysis

1. Introduction

Direct generalization of multivariate techniques to the realm of functional data analysis is not in general feasible, and in this paper some procedures for the one- and multi-sample problem will be modified so as to become suitable for functional data. For an extensive discussion of functional data see the monograph by Ramsay and Silverman [28]. In this paper, the problem of identifying the projective shape of a planar curve will be considered as a practical application.

* Corresponding author. Fax: +1 806 742 1112.
E-mail address: r.paige@ttu.edu (R. Paige).

[1] Research supported by Deutsche Forschungsgemeinschaft MU 1230/10-1.

[2] Research supported by National Science Foundation Grant DMS-0652353.

[3] Research supported by National Science Foundation Grant DMS-0203942.

0047-259X/$ - see front matter
doi:10.1016/j.jmva.2007.04.007

The union–intersection principle of Roy and Bose [29] provides us with a projection pursuit type technique to construct multivariate procedures from a family of univariate procedures. A case in point is Hotelling's [17] multivariate T^2-statistic that can be constructed from a family of univariate student statistics. It is easy to see that further extension to infinite dimensional Hilbert spaces along similar lines breaks down, in particular because the rank of the sample covariance operator cannot exceed the finite sample size and consequently cannot be injective, not even when the population covariance operator is one-to-one.

Several alternatives could be considered. One possibility is projection of the data onto a Euclidean subspace of sufficiently high dimension and perform a Hotelling test with these finite dimensional data. This includes spectral-cut-off regularization of the inverse of the sample covariance operator as a special case. Another option is a Moore–Penrose type of regularization of this operator.

For the application to shape analysis to be considered in this paper, however, yet another modification seems more appropriate. This modification yields at once more realistic hypothesis and a mathematically tractable procedure. In practice "equality of shapes" will almost always refer to a satisfactory visual resemblance rather than exact correspondence in every minute detail. Therefore in this paper the usual hypothesis will be replaced with a "neighborhood hypothesis".

This kind of modified hypothesis has a long history and has been developed in different situations. It has been, e.g. proposed by Hodges and Lehmann [16] for testing whether multinomial cell probabilities are approximately equal. Dette and Munk [9] have extended this approach for the purpose of validating a model in a nonparametric regression framework. For methodological aspects and a more recent discussion we refer to Goutis and Robert [14], Dette and Munk [10], and Liu and Lindsay [21]. The underlying idea is that the hypothesis is often formulated on the basis of theoretical considerations that will never cover reality completely. Hence in practice such a hypothesis will always be rejected if the sample size is large enough. It is therefore more realistic to test a somewhat larger hypothesis that also includes parameters in a neighborhood of the original one. See also Berger and Delampady [1] who employ the term "precise hypothesis" instead of "neighborhood hypothesis", whereas Liu and Lindsay [21] coined the phrase "tubular models". Mostly related to the present approach is the work of Dette and Munk [9] and Munk and Dette [25] who consider L^2-neighborhood hypotheses in nonparametric regression models.

A further advantage is that neighborhood hypotheses often lead to simpler asymptotic analyses. This in turn makes it possible to interchange the role of a neighborhood hypothesis and its alternative without complicating the testing procedure. This is particularly relevant for goodness-of-fit type tests, where traditionally the choice of the null hypothesis is usually dictated by mathematical limitations rather than statistical considerations. Accepting a model after a goodness-of-fit test always leaves the statistician in the ambiguous situation whether the model has not been rejected by other reasons, e.g. because of lack of data, an inefficient goodness-of-fit test at hand, or because of a large variability of the data. In contrast, the present approach allows one to validate a hypotheses at a given level α, instead of accepting a model without any further evidence in favor of the model. In fact, this is equivalent to reporting on a confidence interval for a certain distance measure between models.

There is an objective, data-driven method to select the parameter δ, say, that determines the size of the neighborhood hypothesis. Given any level $\alpha \in (0, 1)$ for the test, one might determine the smallest value $\hat{\delta}(\alpha)$ for which the neighborhood hypothesis is not rejected. It should be realized that modification of Hotelling's test will require a more or less arbitrary regularization parameter.

The paper is organized as follows. In Section 2 we briefly review some basic concepts for Hilbert space valued random variables, and in Section 3 we briefly discuss the difficulties with studentization in infinite dimensional Hilbert spaces. Sections 4 and 5 are devoted, respectively, to a suitably formulated version of the functional one- and multi-sample problem. The theory is applied to the recognition of the projective shape of a planar curve in Sections 6 and 7.

2. Random elements in Hilbert spaces

Let $(\Omega, \mathcal{W}, \mathbb{P})$ be an underlying probability space, $\mathbb{H}$ a separable Hilbert space over the real numbers with inner product $\langle \bullet, \bullet \rangle$ and norm $\| \bullet \|$, and $\mathcal{B}_{\mathbb{H}}$ the σ-field generated by the open subsets of $\mathbb{H}$. A random element in $\mathbb{H}$ is a mapping $X : \Omega \to \mathbb{H}$ which is $(\mathcal{W}, \mathcal{B}_{\mathbb{H}})$-measurable. Let us write $\mathbb{P}_X = P$ for the induced probability measure on $(\mathbb{H}, \mathcal{B}_{\mathbb{H}})$.

The probability distribution P is uniquely determined by its characteristic functional

$$\tilde{P}(x) = \mathbb{E}e^{i\langle x, X\rangle} = \int_{\mathbb{H}} e^{i\langle x, y\rangle}\, dP(y), \quad x \in \mathbb{H}. \tag{2.1}$$

Assuming that

$$E\|X\|^2 < \infty, \tag{2.2}$$

the Riesz representation theorem ensures the existence of a vector $\mu \in \mathbb{H}$ and an operator $\tilde{\Sigma} : \mathbb{H} \to \mathbb{H}$, uniquely determined by the properties

$$\mathbb{E}\langle x, X\rangle = \langle x, \mu\rangle \quad \forall x \in \mathbb{H}, \tag{2.3}$$

$$\mathbb{E}\langle x, X - \mu\rangle \langle y, X - \mu\rangle = \left\langle x, \tilde{\Sigma} y\right\rangle \quad \forall x, y \in \mathbb{H}. \tag{2.4}$$

The operator $\tilde{\Sigma}$ is linear, Hermitian, semi-definite positive; it has, moreover, finite trace and is consequently compact. Any operator with these properties will be referred to as a covariance operator, and any covariance operator is induced by some random element.

It follows from the Minlos–Sazanov theorem that for $\mu \in \mathbb{H}$ and $\tilde{\Sigma} : \mathbb{H} \to \mathbb{H}$ a covariance operator, the functional

$$\varphi(x) = e^{i\langle x, \mu\rangle - \frac{1}{2}\left\langle x, \tilde{\Sigma} x\right\rangle}, \quad x \in \mathbb{H} \tag{2.5}$$

is the characteristic functional of a probability measure on $\mathbb{H}$, which is called the Gaussian measure with parameters μ and $\tilde{\Sigma}$ and will be denoted by $\mathcal{G}(\mu, \tilde{\Sigma})$. The parameters represent, respectively, the mean and covariance operator of the distribution.

Let $\mathbb{H}^p$ be the real, separable Hilbert space of all p-tuples $x = (x_1, \ldots, x_p)^*$, $x_j \in \mathbb{H}$ for $j = 1, \ldots, p$. The inner product in $\mathbb{H}^p$ is given by $\langle x, y\rangle_p = \sum_{j=1}^{p} \left\langle x_j, y_j\right\rangle$, for $x, y \in \mathbb{H}^p$.

3. Why studentization breaks down in $\mathbb{H}$

Let $X_1, \ldots, X_n$ be independent copies of a random element X in $\mathbb{H}$ with

$$\mathbb{E}\|X\|^4 < \infty, \tag{3.1}$$

mean $\mu \in \mathbb{H}$, and covariance operator $\tilde{\Sigma} : \mathbb{H} \to \mathbb{H}$. Estimators of μ and $\tilde{\Sigma}$ are, respectively,

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i, \quad S = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X}) \otimes (X_i - \overline{X}), \tag{3.2}$$

where for $a, b \in \mathbb{H}$ the operator $a \otimes b : \mathbb{H} \to \mathbb{H}$ is defined by $(a \otimes b)(x) = \langle b, x \rangle\, a$, $x \in \mathbb{H}$.

Immediate extension of the union–intersection principle would suggest to use the Hotelling-type test statistic

$$T_n^2 = n \sup_{u \in \mathbb{H}: \|u\|=1} \frac{\left\langle \overline{X}, u \right\rangle^2}{\langle u, Su \rangle}, \tag{3.3}$$

for testing the classical hypothesis that $\mu = 0$. The studentization, however, now in general causes a problem since under the assumption that

$$\mathbb{P}\{X_1, \ldots, X_n \text{ are linearly independent}\} = 1, \tag{3.4}$$

it will be shown that

$$\mathbb{P}\{T_n^2 = \infty\} = 1, \tag{3.5}$$

even when $\tilde{\Sigma}$ is supposed to be injective.

To prove (3.5) let us first observe that (3.4) entails that $\mathbb{P}\{\overline{X} \in$ linear span of $X_1 - \overline{X}, \ldots, X_n - \overline{X}\} = 0$. For if $\overline{X}$ were an element of the linear span there would exist scalars $\alpha_1, \ldots, \alpha_n$ such that $\overline{X} = \sum_{i=1}^{n} \alpha_i (X_i - \overline{X})$. Because of the linear independence of the X_i this means that the vector $\alpha = (\alpha_1, \ldots, \alpha_n)^* \in \mathbb{R}^n$ must satisfy

$$\left(I_n - \frac{1}{n} 1_n 1_n^* \right) \alpha = 1_n, \tag{3.6}$$

where I_n is the $n \times n$ identity matrix and 1_n a column of n numbers 1. This is impossible because the matrix on the left in (3.6) is the projection onto the orthogonal complement in $\mathbb{R}^n$ of the line through 1_n. Hence with probability 1 there exist $\overline{X}_1, \overline{X}_2$ such that $\overline{X} = \overline{X}_1 + \overline{X}_2$, and

$$\begin{cases} \overline{X}_1 \neq 0, \overline{X}_1 \perp X_i - \overline{X} \quad \text{for } i = 1, \ldots, n, \\ \overline{X}_2 \in \text{ linear span of } X_1 - \overline{X}, \ldots, X_n - \overline{X}. \end{cases} \tag{3.7}$$

Choosing $u = \overline{X}_1$ we have on the one hand that $\left\langle \overline{X}, \overline{X}_1 \right\rangle^2 = \|\overline{X}_1\|^4 > 0$, and on the other hand we have $S\overline{X}_1 = n^{-1} . \sum_{i=1}^{n} \left\langle X_i - \overline{X}, \overline{X}_1 \right\rangle (X_i - \overline{X}) = 0$, so that (3.5) follows.

A possible modification of this statistic is obtained by replacing S^{-1} with a regularized inverse of Moore–Penrose type and by considering

$$\begin{aligned} &\sup_{u \in \mathbb{H}: \|u\|=1} \frac{\left\langle \overline{X}, u \right\rangle^2}{\left\langle u, (\alpha I + S)^{-1} u \right\rangle} \\ &\quad = \text{largest eigenvalue of } (\alpha I + S)^{-1/2} \left(\overline{X} \otimes \overline{X} \right) (\alpha I + S)^{-1/2}, \end{aligned}$$

where I is the identity operator. We conjecture that perturbation theory for compact operators in Hilbert spaces leads to the asymptotic distribution of $(\alpha I + S)^{-1/2}$ and subsequently to the asymptotic distribution of this largest eigenvalue, in the same vein as this kind of result can be obtained for matrices. See, for instance, Watson [35] for sample covariance matrices and

Ruymgaart and Yang [30] for functions of sample covariance matrices. Watson's [35] result has been obtained for sample covariance operators on Hilbert spaces by Dauxois et al. [8]. As has been explained in the Introduction, however, here we prefer to pursue the approach of modifying the hypothesis.

4. The one-sample problem in $\mathbb{H}$

Let $X_1, \ldots, X_n$ be as defined in Section 3 and suppose we want to test hypotheses regarding μ. This modified hypothesis may make more sense from an applied point of view and leads, moreover, to simpler asymptotics. To describe these hypotheses suppose that

$$M \subset \mathbb{H} \text{ is a linear subspace of dimension } m \in \mathbb{N}_0, \tag{4.1}$$

and let $\delta > 0$ be an arbitrary given number. Let us denote the orthogonal projection onto M by Π, and onto $M^\perp$ by $\Pi^\perp$. It is useful to observe that

$$\left\langle \Pi^\perp x, \Pi^\perp y \right\rangle = \left\langle x, \Pi^\perp y \right\rangle \quad \forall x, y \in \mathbb{H}. \tag{4.2}$$

Furthermore let us introduce the functional

$$\varphi_M(x) = \|x - M\|^2 = \left\| \Pi^\perp x \right\|^2, \quad x \in \mathbb{H}, \tag{4.3}$$

representing the squared distance of a point $x \in \mathbb{H}$ to M (finite dimensional subspaces are closed).

The "neighborhood hypothesis" to be tested is

$$\mathcal{H}_\delta : \mu \in M_\delta \cup B_\delta \quad \text{for some } \delta > 0, \tag{4.4}$$

where $M_\delta = \{x \in \mathbb{H} : \varphi_M(x) < \delta^2\}$ and $B_\delta = \{x \in \mathbb{H} : \varphi_M(x) = \delta^2, \left\langle \Pi^\perp x, \tilde{\Sigma}\Pi^\perp x \right\rangle > 0\}$. The alternative to (4.4) is

$$\mathcal{A}_\delta : \mu \in M_\delta^c \cap B_\delta^c. \tag{4.5}$$

The usual hypothesis would have been: $\mu \in M$. It should be noted that $\mathcal{H}_\delta$ contains $\{\varphi_M < \delta^2\}$ and that $\mathcal{A}_\delta$ contains $\{\varphi_M > \delta^2\}$. These are the important components of the hypotheses; the set B_δ is added to the null hypothesis by mathematical convenience, i.e. because the asymptotic power on that set is precisely α, as will be seen below.

For testing hypotheses like (4.4) see Dette and Munk [9]. These authors also observe that testing

$$\mathcal{H}'_\delta : \mu \in \left(M'_\delta\right)^c \cup B_\delta \quad \text{versus} \quad \mathcal{A}'_\delta : \mu \in M'_\delta \cap B_\delta^c, \tag{4.6}$$

where $M'_\delta = \{x \in \mathbb{H} : \varphi_M(x) > \delta^2\}$, can be done in essentially the same manner; see also Remark 1. This may be very useful in practice. When, for instance, M is the subspace of all polynomials of degree at most $m-1$, it is more appropriate to test (4.6) if one wants to establish that the mean value function is close to such a polynomial. In the traditional set-up interchanging null hypothesis and alternative would be virtually impossible due to mathematical difficulties, just as this is the case in the classical goodness-of-fit problems.

The reason that it is mathematically easier to deal with the present hypotheses is that the test statistic, which is based on

$$\varphi_M(\overline{X}) - \delta^2 \tag{4.7}$$

has a simple normal distribution in the limit for large sample sizes.

Lemma 1. *We have*

$$\sqrt{n}\{\varphi_M(\overline{X}) - \varphi_M(\mu)\} \to_d \mathcal{N}(0, v^2) \quad \text{as } n \to \infty, \tag{4.8}$$

where

$$v^2 = 4\left\langle \Pi^\perp \mu, \tilde{\Sigma}\Pi^\perp \mu \right\rangle. \tag{4.9}$$

If $v^2 = 0$ *the limiting distribution* $\mathcal{N}(0, 0)$ *is to be interpreted as the distribution which is a degenerate at* 0.

Proof. The central limit theorem for $\mathbb{H}$-valued random variables yields the existence of a $\mathcal{G}(0, \tilde{\Sigma})$ random element G, such that

$$\sqrt{n}(\overline{X} - \mu) \to_d G \quad \text{as } n \to \infty, \tag{4.10}$$

in $(\mathbb{H}, \mathcal{B}_{\mathbb{H}})$. It is easy to see that $\varphi_M : \mathbb{H} \to \mathbb{R}$ is Fréchet differentiable at any $\mu \in \mathbb{H}$, tangentially to $\mathbb{H}$, with derivative the linear functional

$$2\left\langle \Pi^\perp \mu, h\right\rangle, \quad h \in \mathbb{H}. \tag{4.11}$$

According to the "functional delta method" we may conclude

$$\sqrt{n}\{\varphi_M(\overline{X}) - \varphi_M(\mu)\} \to_d 2\left\langle \Pi^\perp \mu, G\right\rangle. \tag{4.12}$$

The random variable on the right in (4.12) is normal, because G is Gaussian, and clearly its mean is 0. Therefore its variance equals

$$\mathbb{E}\left\langle \Pi^\perp \mu, G\right\rangle\left\langle \Pi^\perp \mu, G\right\rangle = \left\langle \Pi^\perp \mu, \tilde{\Sigma}\Pi^\perp \mu\right\rangle, \tag{4.13}$$

according to the definition of $\tilde{\Sigma}$ (cf. (2.4)). □

Lemma 2. *We have*

$$\hat{v}_n^2 = 4\left\langle \Pi^\perp \overline{X}, S\Pi^\perp \overline{X}\right\rangle \to_p v^2 \quad \text{as } n \to \infty. \tag{4.14}$$

Proof. By simple algebra we find

$$\begin{aligned}
\left\langle \Pi^\perp \overline{X}, S\Pi^\perp \overline{X}\right\rangle &= \left\langle \Pi^\perp \overline{X}, \frac{1}{n}\sum_{i=1}^n \left\langle X_i - \overline{X}, \Pi^\perp \overline{X}\right\rangle (X_i - \overline{X})\right\rangle \\
&= \frac{1}{n}\sum_{i=1}^n \left\langle X_i - \overline{X}, \Pi^\perp \overline{X}\right\rangle^2 \\
&= \frac{1}{n}\sum_{i=1}^n \left\{\left\langle X_i - \mu, \Pi^\perp \mu\right\rangle + \left\langle X_i - \mu, \Pi^\perp(\overline{X} - \mu)\right\rangle\right. \\
&\quad \left. + \left\langle \mu - \overline{X}, \Pi^\perp \mu\right\rangle + \left\langle \mu - \overline{X}, \Pi^\perp(\overline{X} - \mu)\right\rangle\right\}^2.
\end{aligned} \tag{4.15}$$

According to the weak law of large numbers and the definition of covariance operator we have

$$\frac{1}{n}\sum_{i=1}^n \left\langle X_i - \mu, \Pi^\perp \mu\right\rangle^2 \to_p \mathbb{E}\left\langle X - \mu, \Pi^\perp \mu\right\rangle^2 = \left\langle \Pi^\perp \mu, \tilde{\Sigma}\Pi^\perp \mu\right\rangle \quad \text{as } n \to \infty.$$

All the other terms tend to 0 in probability. As an example consider

$$\left\langle \mu - \overline{X}, \Pi^{\perp}\mu \right\rangle^2 \leqslant \|\overline{X} - \mu\|^2 \|\Pi^{\perp}\mu\|^2 \rightarrow_p 0,$$

as $n \to \infty$. The lemma follows from straightforward combination of the above. $\square$

For $0 < \alpha < 1$ let $\xi_{1-\alpha}$ denote the quantile of order $1-\alpha$ of the standard normal distribution. Focusing on the testing problem (4.4), (4.5) let us decide to reject the null hypothesis when $\sqrt{n}\{\varphi_M(\overline{X}) - \delta^2\}/\hat{v} > \xi_{1-\alpha}$. The corresponding power function is then

$$\beta_n(\mu) = \mathbb{P}\{\sqrt{n}\{\varphi_M(\overline{X}) - \delta^2\}/\hat{v} > \xi_{1-\alpha}\}, \tag{4.16}$$

when $\mu \in \mathbb{H}$ is the true parameter.

Theorem 4.1. *Asymptotics under the null hypothesis and fixed alternatives. The power function in* (4.16) *satisfies*

$$\lim_{n\to\infty} \beta_n(\mu) = \begin{cases} 0, & \varphi_M(\mu) < \delta^2, \\ \alpha, & \varphi_M(\mu) = \delta^2, v^2 > 0, \\ 1, & \varphi_M(\mu) > \delta^2. \end{cases} \tag{4.17}$$

Hence the test has asymptotic size α, *and is consistent against the alternatives* $\mu : \varphi_M(\mu) > \delta^2$.

Proof. If $v^2 > 0$ it is immediate from Lemmas 1 and 2 that $\sqrt{n}\{\varphi_M(\overline{X}) - \delta^2\}/\hat{v} \longrightarrow_d \mathcal{N}(0, 1)$. The result now follows in the usual way by observing that $\sqrt{n}\{\delta^2 - \varphi_M(\mu)\}$ tends to either ∞ (when $\varphi_M(\mu) < \delta^2$), to 0 (when $\varphi_M(\mu) = \delta^2$) or to $-\infty$ (when $\varphi_M(\mu) > \delta^2$). If $v^2 = 0$ we still have that $\sqrt{n}\{\varphi_M(\overline{X}) - \delta^2\}/\hat{v}$ tends in probability to ∞ (when $\varphi_M(\mu) < \delta^2$) or to $-\infty$ (when $\varphi_M(\mu) > \delta^2$). $\square$

To describe the sampling situation under local alternatives (including the null hypothesis) we assume now that

$$X_1, \ldots, X_n \text{ are i.i.d. } (\mu_{n,t}, \tilde{\Sigma}), \tag{4.18}$$

where $\tilde{\Sigma}$ is as above and

$$\mu_{n,t} = \mu + \frac{t}{\sqrt{n}}\gamma, \quad t \geqslant 0, \tag{4.19}$$

for some (cf. (4.4) and below)

$$\mu \in B_\delta, \quad \gamma \in \mathbb{H} : \left\langle \mu, \Pi^{\perp}\gamma \right\rangle > 0. \tag{4.20}$$

Under these assumptions it follows that $\mu_{n,0} = \mu$ satisfies $\mathcal{H}_\delta$, and $\mu_{n,t}$ satisfies $\mathcal{A}_\delta$ for each $t > 0$. Let Φ denote the standard normal c.d.f.

Theorem 4.2. *Asymptotic power. We have*

$$\lim_{n\to\infty} \beta_n(\mu_{n,t}) = 1 - \Phi\left(\xi_{1-\alpha} - 2t\frac{\langle \mu, \Pi^{\perp}\gamma\rangle}{v}\right), \quad t > 0. \tag{4.21}$$

Proof. We may write $X_i = X_i' + (t/\sqrt{n})\gamma$, where the X_i' are i.i.d. $(\mu, \tilde{\Sigma})$. It is easy to see from this representation that we still have

$$\hat{v}_n^2 \longrightarrow_p v^2 > 0 \quad \text{as } n \to \infty \ \forall t > 0. \tag{4.22}$$

Exploiting once more the Fréchet differentiability of φ_M (see (4.11)) we obtain

$$\sqrt{n}\left\{\frac{\varphi_M(\overline{X}) - \delta^2}{\hat{v}}\right\} = \sqrt{n}\left\{\frac{\varphi_M(\overline{X}') - \varphi(\mu)}{\hat{v}}\right\} + 2t\left\{\frac{\langle \Pi^\perp \mu, \Pi^\perp \gamma\rangle}{\hat{v}}\right\}$$
$$+o_p(1) \to_d \mathcal{N}\left(2t\frac{\langle \mu, \Pi^\perp \gamma\rangle}{v}, 1\right) \quad \text{as } n \to \infty, \tag{4.23}$$

and the result follows. $\square$

Remark 1. To corroborate the remark about interchanging null hypothesis and alternative made at the beginning of this section, just note that an asymptotic size α test for testing $\mathcal{H}_\delta'$ versus $\mathcal{A}_\delta'$ in (4.6) is obtained by rejecting $\mathcal{H}_\delta'$ when

$$\sqrt{n}\left\{\varphi_M(\overline{X}) - \delta^2\right\}/\hat{v} < \xi_\alpha, \quad \alpha \in (0, 1). \tag{4.24}$$

This allows to assess the approximate validity of the model within the neighborhood δ. Of course, from (4.24) we immediately get a confidence interval for δ as well.

Remark 2. The expression in (4.23) remains valid for $t = 0$ or $\gamma = 0$. In either case the corresponding mean satisfies the null hypothesis assumption and the limit in (4.23) equals α.

5. The multi-sample problem in $\mathbb{H}$

Let $X_{j1}, \ldots, X_{jn_j}$ be i.i.d. with mean μ_j and covariance operator $\tilde{\Sigma}_j$, where $n_j \in \mathbb{N}$, s.t. $\sum_j n_j = n$, and let these random elements satisfy the moment condition in (3.1): all of this for $j = 1, \ldots, p$. Moreover these p samples are supposed to be mutually independent, and their sample sizes satisfy

$$\begin{cases} \dfrac{n_j}{n} = \lambda_j + o\left(\dfrac{1}{\sqrt{n}}\right) & \text{as } n = n_1 + \cdots + n_p \to \infty, \\ \lambda_j \in (0, 1), \quad j = 1, \ldots, p. \end{cases} \tag{5.1}$$

Let us define

$$\overline{X}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} X_{ji}, \quad \overline{X} = \frac{1}{p}\sum_{j=1}^{p}\frac{n_j}{n}\overline{X}_j, \quad j = 1, \ldots, p. \tag{5.2}$$

Furthermore, let the functionals $\psi_n : \mathbb{H}^p \to \mathbb{R}$ be given by

$$\psi_n(x_1, \ldots, x_p) = \sum_{j=1}^{p}\left\|\frac{n_j}{n}x_j - \overline{x}_n\right\|^2, \tag{5.3}$$

where $x_1, \ldots, x_p \in \mathbb{H}$ and $\overline{x}_n = \frac{1}{p}\sum_{j=1}^{p} \frac{n_j}{n} x_j$. Defining $\psi : \mathbb{H}^p \to \mathbb{R}$ by

$$\psi(x_1, \ldots, x_p) = \sum_{j=1}^{p} \left\| \lambda_j x_j - \overline{x} \right\|^2, \tag{5.4}$$

where $\overline{x} = \frac{1}{p}\sum_{j=1}^{p} \lambda_j x_j$, it is readily verified that

$$\sqrt{n}\{\psi_n(x_1, \ldots, x_p) - \psi(x_1, \ldots, x_p)\} \to 0 \quad \text{as } n \to \infty, \tag{5.5}$$

provided that condition (5.1) is fulfilled.

The neighborhood hypothesis in this model can be loosely formulated as "approximate equality of the means". More precisely the null hypothesis

$$\mathcal{H}_{p,\delta} : \mu = (\mu_1, \ldots, \mu_p)^* \in M_{p,\delta} \cup B_{p,\delta}, \tag{5.6}$$

where $M_{p,\delta} = \{x \in \mathbb{H}^p : \psi(x) < \delta^2\}$ and $B_{p,\delta} = \{x \in \mathbb{H}^p : \psi(x) = \delta^2, \sum_{j=1}^{p} \lambda_j \left\langle \lambda_j x_j - \overline{x}, \tilde{\Sigma}_j(\lambda_j x_j - \overline{x}) \right\rangle > 0\}$, will be tested against the alternative

$$\mathcal{A}_{p,\delta} : \mu = (\mu_1, \ldots, \mu_p)^* \in M_{p,\delta}^c \cap B_{p,\delta}^c. \tag{5.7}$$

Let us introduce some further notation and set

$$\tau_p^2 = 4\sum_{j=1}^{p} \lambda_j \left\langle \lambda_j \mu_j - \overline{\mu}, \tilde{\Sigma}_j(\lambda_j \mu_j - \overline{\mu}) \right\rangle, \quad \overline{\mu} = \frac{1}{p}\sum_{j=1}^{p} \lambda_j \mu_j. \tag{5.8}$$

Writing S_j for the sample covariance operator of the jth sample (cf. (3.2)) the quantity in (5.8) will be estimated by

$$\hat{\tau}_{n,p}^2 = 4\sum_{j=1}^{p} \lambda_j \left\langle \lambda_j \overline{X}_j - \overline{X}, S_j(\lambda_j \overline{X}_j - \overline{X}) \right\rangle. \tag{5.9}$$

Theorem 5.1. *The test that rejects $\mathcal{H}_{p,\delta}$ for*

$$\sqrt{n}\left\{\psi_n(\overline{X}_1, \ldots, \overline{X}_p) - \psi_n(\mu_1, \ldots, \mu_p)\right\}/\hat{\tau}_{p,n} > \xi_{1-\alpha}, \quad 0 < \alpha < 1 \tag{5.10}$$

has asymptotic size α, and is consistent against fixed alternatives $\mu = (\mu_1, \ldots, \mu_p)^*$ with $\psi(\mu) > \delta^2$.

Proof. Because the p samples are independent the central limit theorem in (4.10) yields

$$\sqrt{n}\begin{pmatrix} \overline{X}_1 - \mu_1 \\ \vdots \\ \overline{X}_p - \mu_p \end{pmatrix} \to_d \begin{pmatrix} G_1 \\ \vdots \\ G_p \end{pmatrix}, \tag{5.11}$$

where $G_1, \ldots, G_p$ are independent Gaussian random elements in $\mathbb{H}$, and

$$G_j =_d \mathcal{G}\left(0, \frac{1}{\lambda_j}\tilde{\Sigma}_j\right). \tag{5.12}$$

It follows from (5.5) that

$$\sqrt{n}[\psi_n(\overline{X}_1,\ldots,\overline{X}_p)-\psi_n(\mu_1,\ldots,\mu_p)-\{\psi(\overline{X}_1,\ldots,\overline{X}_p)-\psi(\mu_1,\ldots,\mu_p)\}]$$
$$= o_p(1). \tag{5.13}$$

Moreover, a slight modification of Lemma 2 yields that $\left\langle \overline{X}_j-\overline{X}, S_j(\overline{X}_j-\overline{X})\right\rangle \to_p \left\langle \mu_j - \overline{\mu}, \sum_j(\mu_j-\overline{\mu})\right\rangle$ and hence

$$\hat{\tau}^2_{n,p} \to_p \tau^2_p. \tag{5.14}$$

This means that the statistic on the left in (5.10) and the one obtained by replacing ψ_n with ψ in that expression will exhibit the same first order asymptotics. The proof will be continued with the latter, simpler version. A simple calculation shows that $\psi : \mathbb{H}^p \to \mathbb{R}$ is Fréchet differentiable at any $x \in \mathbb{H}^p$, tangentially to $\mathbb{H}^p$. Writing $\overline{h} = \frac{1}{p}\sum_{j=1}^p \lambda_j h_j$, for any $h_1,\ldots,h_p \in \mathbb{H}$, its derivative is equal to

$$2\sum_{j=1}^{p}\left\langle \lambda_j x_j-\overline{x}, \lambda_j h_j-\overline{h}\right\rangle = 2\sum_{j=1}^{p}\left\langle \lambda_j x_j-\overline{x}, \lambda_j h_j\right\rangle. \tag{5.15}$$

Application of the delta method with the functional ψ in the basic result (5.11) yields

$$\sqrt{n}\{\psi(\overline{X}_1,\ldots,\overline{X}_p)-\psi(\mu_1,\ldots,\mu_p)\} \to_d 2\sum_{j=1}^{p}\left\langle \lambda_j\mu_j-\overline{\mu}, \lambda_j G_j\right\rangle. \tag{5.16}$$

According to (5.12) we have

$$\lambda_j G_j =_d \mathcal{G}\left(0, \lambda_j\tilde{\Sigma}_j\right), \tag{5.17}$$

and because of the independence of the G_j it follows that

$$2\sum_{j=1}^{p}\left\langle \lambda_j\mu_j-\overline{\mu}, \lambda_j G_j\right\rangle =_d \mathcal{N}(0,\tau^2_p), \tag{5.18}$$

where τ^2_p is defined in (5.8). Exploiting the consistency in (5.14) the proof can be concluded in much the same way as that of Theorem 4.1. Just as in that theorem we need here that $\tau^2_p > 0$ at the alternative considered in order to ensure consistency. □

6. Hilbert space representations of projective shapes of planar curves

A nonsingular matrix $A = (a^j_i)_{i,j=0,\ldots,m}$ defines a projective transformation in $\mathbb{R}^m$ given by

$$(y^1,\ldots,y^m) = f(x^1,\ldots,x^m),$$
$$y^j = \frac{\sum_{i=0}^m a^j_i x^i}{\sum_i a^0_i x^i}, \forall j = 1,\ldots,m. \tag{6.1}$$

Two configurations of points in $\mathbb{R}^m$ have the same *projective shape* if they differ by a projective transformation of $\mathbb{R}^m$. Unlike similarities or affine transformations, projective transformations do not have a group structure under composition, since the domain of definition of a projective transformation depends on the transformation, and the maximal domain of a composition has to be restricted accordingly. To avoid such unwanted situations, rather than considering projective shapes of configurations in $\mathbb{R}^m$, one may consider configurations in the real projective space $\mathbb{R}P^m$, with the projective general linear group action that is described below. We recall that the real projective space in m dimensions, $\mathbb{R}P^m$, is the set of axes going through the origin of $\mathbb{R}^{m+1}$. If $u = (u^1, \dots, u^{m+1}) \in \mathbb{R}^{m+1}\backslash\{0\}$, then

$$[u] = [u^1 : u^2 : \dots : u^{m+1}] = \{\lambda u, \lambda \neq 0\} \tag{6.2}$$

is a (*projective*) *point* in $\mathbb{R}P^m$; the notation $[\cdot]$ for the projective points will be used throughout. The affine space $\mathbb{R}^m$ is canonically embedded in $\mathbb{R}P^m$ via

$$(x^1, \dots, x^m) \longrightarrow [x^1 : \dots : x^m : 1]. \tag{6.3}$$

Via this embedding, the pseudogroup action of projective transformations (6.1) on $\mathbb{R}^m$ can be regarded as a restriction of the action of the *projective general linear group* $PGL(m)$ on $\mathbb{R}P^m$; this group and its action on $\mathbb{R}P^m$ are defined in terms of the natural action of $GL(m+1)$ on $\mathbb{R}^{m+1}$ as defined below. To each matrix $A \in GL(m+1)$ we associate an element $\alpha \in PGL(m)$ whose action on $\mathbb{R}P^m$ is given by

$$[u'] = \alpha([u]) = [Au]. \tag{6.4}$$

By way of example, let us consider the set $G_0(k, m)$ of k projective points $(p_1, \dots, p_k)$, $k \geqslant m+2$ for which $(p_1, \dots, p_{m+2})$ is a projective frame in $\mathbb{R}P^m$. $PGL(m)$ acts diagonally on $G_0(k, m)$ by $\alpha(p_1, \dots, p_k) = (\alpha(p_1), \dots, \alpha(p_k))$. $P\Sigma_m^k$, space of orbits of k-tuples in $\mathbb{R}P^m$ under this action, is the *projective shape space* of k-ads in general position considered in Mardia and Patrangenaru [24].

The projective shape of a configuration made of a projective frame plus an infinite set of projective points can be also represented as a space of configurations in $\mathbb{R}P^m$. Recall that a *projective frame* in $\mathbb{R}P^m$ is an ordered $(m+2)$-tuple of points $\pi = (p_1, \dots, p_{m+2})$, any $m+1$ of which are in general position. The *standard* projective frame π_0 is the projective frame associated with the standard vector basis of $\mathbb{R}^{m+1}$, in this case $p_1 = [e_1], \dots, p_{m+1} = [e_{m+1}]$, $p_{m+2} = [e_1 + \dots + e_{m+1}]$. The action of a projective transformation is uniquely determined by its action on a projective frame. Given a point $p \in \mathbb{R}P^m$ its *projective coordinate* p^π w.r.t. a projective frame $\pi = (p_1, \dots, p_{m+2})$, is the image of p, under the projective transformation that takes π to π_0. The case $m = 2$ is illustrated in Figs. 1 and 2.

In Fig. 1 the round dots yield a projective frame and in Fig. 2 the square dot gives an affine representative of the projective coordinates of the square dot in Fig. 1 with respect to that frame.

Our approach to projective shapes of planar closed curves is based on the idea above of registration with respect to a projective frame. To keep things simple, assume that in addition to a closed planar curve, four labelled control points, that yield a projective frame are also known. Such a configuration will be called *framed closed curve*. Two framed closed curves have the same projective shape if they differ by a planar projective transformation that brings the projective frame in the first configuration into coincidence with the projective frame in the second configuration.

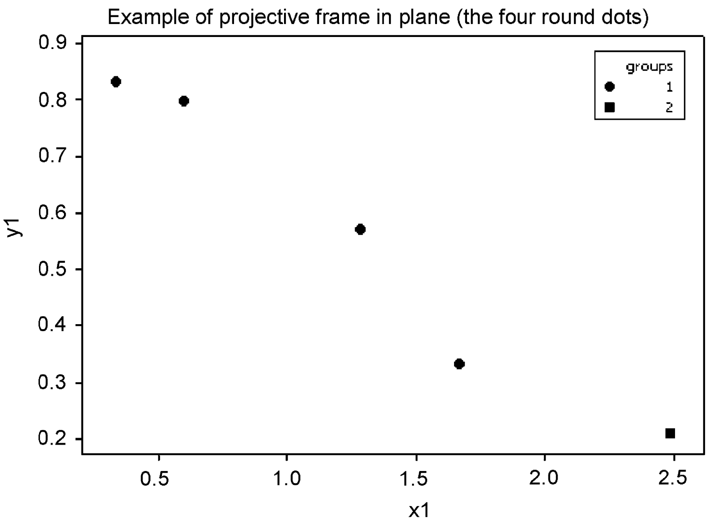

Fig. 1. Affine view of a projective frame in 2D.

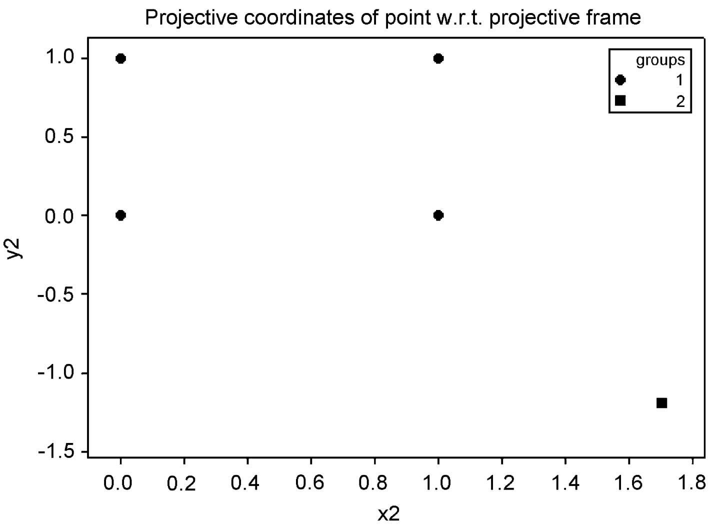

Fig. 2. Affine view of projective coordinates of a point.

Remark 3. In the context of scene recognition, the frame assumption is natural, given that a scene pictured may contain more information than just a curved contour. Such information may include feature landmarks that can be spotted in different images of the scene.

Assume $x_1, \ldots, x_{m+2}$ are points in general position and $x = (x^1, \ldots, x^m)$ is an arbitrary point in $\mathbb{R}^m$. Note that in our notation, the superscripts are reserved for the components of a point, whereas the subscripts are for the labels of points. In order to determine the projective coordinates of $p = [x : 1]$ w.r.t. the projective frame associated with $(x_1, \dots, x_{m+2})$ we set $\tilde{x} = (x^1, \dots, x^m, 1)^T$ and consider the $(m+1) \times (m+1)$ matrix $U_m = [\tilde{x}_1, \dots, \tilde{x}_{m+1}]$, those jth column is $\tilde{x}_j = (x_j, 1)^T$, $j = 1, \dots, m+1$. We define an intermediate system of homogeneous coordinates

$$v(x) = U_m^{-1}\tilde{x} \tag{6.5}$$

and write $v(x) = (v^1(x), \dots, v^{m+1}(x))^T$. Next we set

$$z^j(x) = \frac{v^j(x)}{v^j(x_{m+2})} \Big/ \left\| \frac{v^j(x)}{v^j(x_{m+2})} \right\|, \quad j = 1, \dots, m+1 \tag{6.6}$$

so that the last point x_{m+2} is now used. The projective coordinate(s) of x are given by the point $[z^1(x) : \dots : z^{m+1}(x)]$, where $(z^1(x))^2 + \cdots + (z^{m+}(x))^2 = 1$. If $z^{m+1}(x) \neq 0$, the affine representative of this point with respect to the last coordinate is $(\xi^1(x), \dots, \xi^m(x))$, where

$$\xi^j(x) = \frac{z^j(x)}{z^{m+1}(x)}, \quad j = 1, \dots, m. \tag{6.7}$$

Assume $x(t), t \in I$ is a curve in $\mathbb{R}^m$, such that $\forall t \in I, z^{m+1}(x(t)) \neq 0$. Such framed curves will be said to be in a *convenient position* relative to the projective frame π associated with $(x_1, \dots, x_{m+2})$.

Theorem 6.1. *There is a one-to-one correspondence between the set of projective shapes of framed curves $x(t), t \in I$ in a convenient position relative to π, and curves in $\mathbb{R}^m$. In this correspondence, framed closed curves in a convenient position relative to π correspond to closed curves in $\mathbb{R}^m$.*

We will use the representation Theorem 6.1 for projective shapes of closed curves in the projective space that avoid a hyperplane; they correspond to closed curves in the Euclidean space. In particular in two dimensions we consider framed closed curves in the planar projective plane, avoiding a projective line. In particular if we assume that the $(x(t), y(t)), t \in [0, 1]$ is a closed planar curve, then $[x(t) : y(t) : 1], t \in [0, 1]$ is such a projective curve, and using a projective frame π we associate with this curve the affine representative $(\xi(t), \eta(t)), t \in [0, 1]$ of its curve of projective coordinates $[x(t) : y(t) : 1]^\pi$, which yield another planar curve. If two curves are obtained from a planar curve viewed from different perspective points, then the associated affine curves are the same. This affine representative of the projective curve of a (closed) curve is used in this paper. Here we are concerned with the recognition of a closed curve

$$\gamma(t) = (\xi(t), \eta(t)), \quad t \in [0, 1], \quad (\xi(0), \eta(0)) = (\xi(1), \eta(1)) \tag{6.8}$$

that is observed with random errors

$$\Gamma(t) = (\xi(t), \eta(t)) + (\varepsilon^X(t), \varepsilon^Y(t)), \quad t \in [0, 1], \tag{6.9}$$

where $\varepsilon^X(t)$ and $\varepsilon^Y(t)$ are stochastic independent error processes, $(\varepsilon^X(0), \varepsilon^Y(0)) = (\varepsilon^X(1), \varepsilon^Y(1))$, so that the observed curve can, for instance, be considered as a random element in the Hilbert space $\mathbb{H} = L^2(S^1, \mathbb{R}^2)$.

The distance in $\mathbb{H} = L^2(S^1, \mathbb{R}^2)$ induces a distance on the space of projective shapes of planar closed framed curves in convenient position, and the Fréchet mean of a random closed curve in this space corresponds to the mean of the corresponding $\mathbb{H}$-valued random variable. As an application of the results obtained in Section 4, we consider the null neighborhood hypothesis $\mathrm{H}_\delta : \mu \in \gamma_0 + B_\delta$, for some $\delta > 0$; in this case the linear subspace M is the trivial subspace, which is the infinite dimensional analog of the classical null hypothesis $\mathrm{H}_0 : \mu = \gamma_0$. The constant $\delta > 0$ in (4.4) is to be determined from the data, being estimated from (4.24), as shown in a concrete example in the next section.

7. The one sample problem for mean projective shapes of planar curves

Motivated by high level image analysis, projective shape analysis of a (possibly infinite) random configuration of points in $\mathbb{R}^m$ is concerned with understanding this configuration modulo projective transformations. As described in Ma et al. [23] well-focused digital camera images may be assumed to have come from an ideal pinhole camera. The *real projective plane* $\mathbb{R}P^2$ is a geometric model of a pinhole camera view. Two digital images of the same planar scene, taken by ideal pinhole cameras, differ by a composition of two central projections in $\mathbb{R}^3$ from an observed plane to a receiving plane (retina, film, etc.). Such a map, turns out to be a projective transformation in $\mathbb{R}^2$. Therefore, as far as pinhole camera acquired images is concerned, projective shape is "the" shape of interest in high level image analysis.

Scientists are looking for new computational algorithms, including statistical methods, to deal with digital imaging libraries. Images of approximately planar scenes are very common, and their need being analyzed in their full complexity. Until today only finite configurations were analyzed, although the actual scenes are more complex, including curves, and regions bounded by these curves. A toy example of such images, from the so-called "BigFoot" data set is displayed below. Such data lead us to considering the space of projective shapes of closed planar curves in Fig. 3.

Remark 4. The *similarity shape of a planar curve* is the orbit of the curve (viewed as a possibly re-parameterized curve) under the group of direct similarities of the plane. The space of closed similarity shapes of planar curves is a Hilbert (infinite dimensional) manifold. Certain statistical aspects have been studied by Srivastava et al. [31] and by Klassen et al. [19]. A general space of projective shapes of planar curves, can be also defined in such a general context, nevertheless a statistical analysis on such a Hilbert manifold like object goes beyond our interest.

Our approach to projective shape analysis based on the idea of Hilbert space representation of the projective shape with respect to a projective frame is summarized in Theorem 6.1. To identify the mean projective shape of a curves, one may now use the statistical testing method for functional data, described in Section 4.

In practice two curves will not have exactly the same shape, even if they should agree according to some theory. In this case therefore, using the neighborhood hypothesis, stating the approximate equality of the shapes of the curves, seems appropriate.

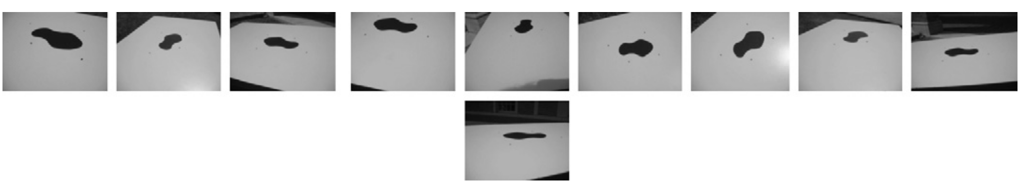

Fig. 3. Ten views of a scene including a natural projective frame (the four points) and a curve (the edge of the footprint).

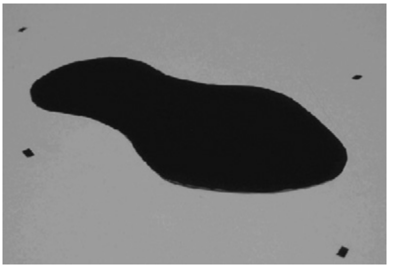

Fig. 4. First “Bigfoot” image.

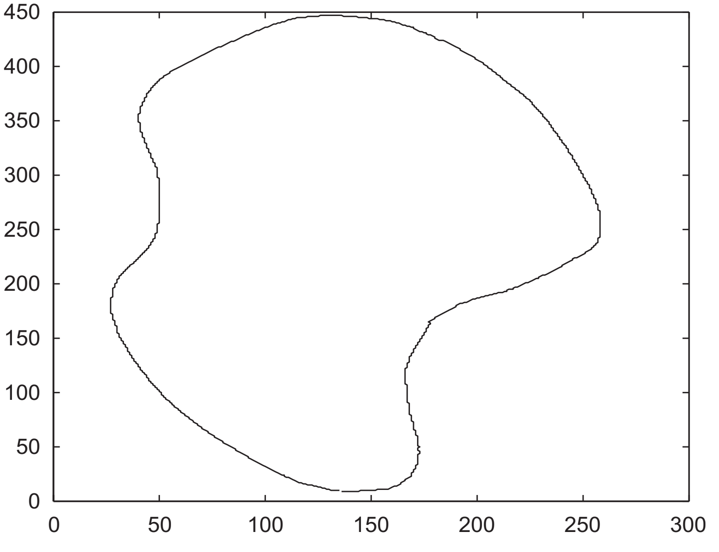

Fig. 5. The processed footprint image.

The implementation of the estimation techniques described in Paige et al. [26], although straightforward, is computationally intensive, given the large number of pixels on a curve, and will be performed in subsequent work. The estimation technique in this section is applied to the “Bigfoot” data set and a new image not necessarily belonging to this data set.

7.1. Image processing and shape registration using the projective frame

Image processing was performed in the Image Processing Toolbox (IPT), MATLAB 7.1. and Microsoft©Paint 5.1. The end result of processing the first “Bigfoot” image shown in Fig. 4 is shown in Fig. 5.

In generating this curve we first crop the original image to remove as much of the noise from image as possible. In this case the “noise” is the grass and the edge of the table. Next, using the projective coordinates w.r.t. the selected projective frame we register the cropped image. After registration, the Sobel method of edge detection is used to extract the edge of the footprint and the

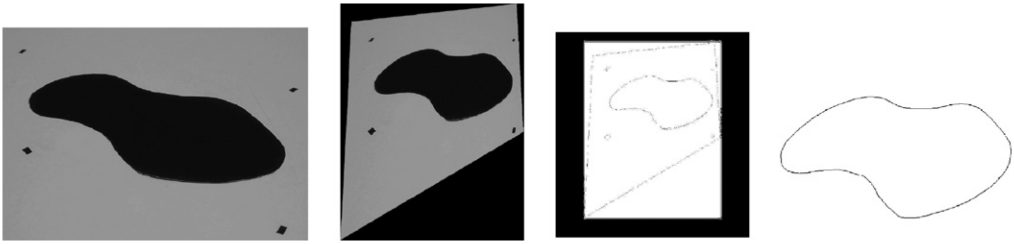

Fig. 6. Registered curve using the projective frame present in original images.

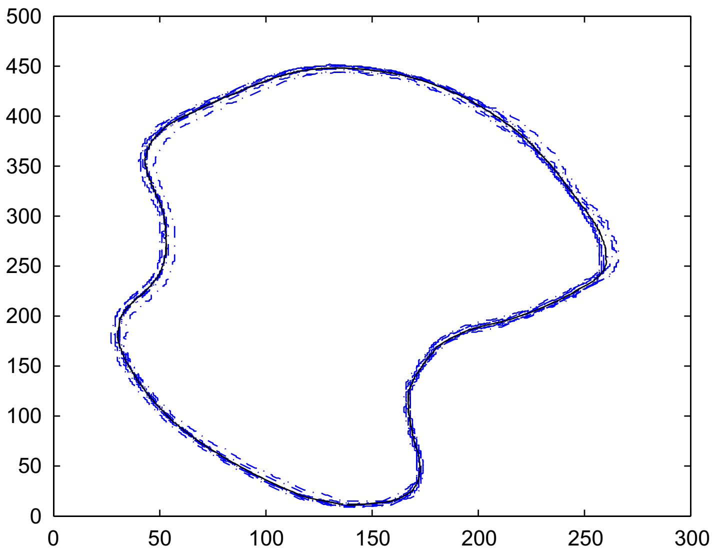

Fig. 7. Empirical sample mean of the observations from the first group.

landmarks. Next, the landmarks and extraneous marks are removed in Microsoft©Paint 5.1 and the image is recropped. Finally, to define a piecewise linear projective curve, pixel locations on curve image are put in clockwise order using MATLAB M-file sort_coord_pixel.m. This shareware M-file was written by Alister Fong and is available for download from the MATLAB Central File Exchange website. A sequence of pictures representing the steps in transforming the image in Fig. 4 into the projective curve in Fig. 5 are displayed in Fig. 6.

Note the ordering of pixel locations in effect rotates the curve image by 90° since the first point on the projective curve corresponds to the first pixel location (moving from top to bottom and from left to right) lying on the curve image. The 10 "Bigfoot" projective curves and their sample mean curve are shown in Fig. 7.

7.2. *Hypothesis testing*

One of the classical problems in pattern recognition, is the identification of a scene for which prior information is known. As a typical example consider that a number of images of a planar

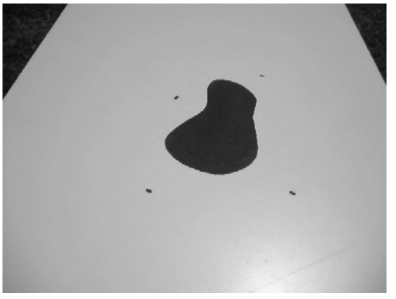

Fig. 8. View of a second unknown scene including a natural projective frame and a curve.

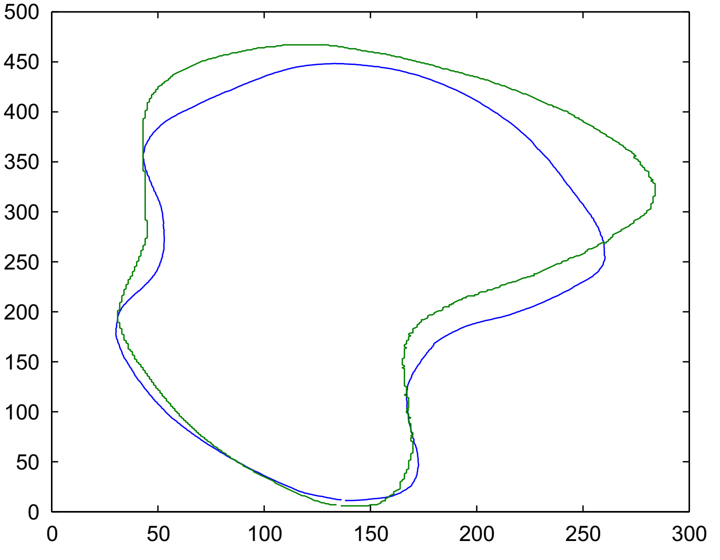

Fig. 9. Plot of the mean "Bigfoot" curve and the new image curve.

scene are known, and we acquire a new image that is apparently of the same scene. In our example, the known data are displayed in Fig. 3 and the new image is shown in Fig. 8.

For this new contour γ_0, we consider the null hypothesis $H_0 : \mu \in \gamma_0 + B_\delta$, for some $\delta > 0$, which is equivalent to (4.4). Testing based upon the asymptotic pivot in (4.24), with $\alpha = 0.05$, yields a δ cutoff value of 312.39. This value represents the largest δ value for which we would reject the null hypothesis.

This means that if we choose $\delta < 312.39$, we would then reject the equality of the mean projective shape of the first population of curves with the projective shape of the second curve, and thus conclude that the mean of the first "Bigfoot" planar scene is significantly different from the projective shape of the curve in Fig. 8.

For a visual understanding of this significant difference, one can compare the curves in Fig. 9. Here the mean curve of the "Bigfoot" sample is plotted along with the new image curve.

Remark 5. The sample size was small, and we could have used nonparametric bootstrap; see Bhattacharya and Denker (1990), Efron (1982) and Patrangenaru (2001). Nevertheless, the errors are quite insignificant, since they depend only on the pose of the scene, which is essentially flat. Thus even for our fairly small sample, the result is reliable.

Remark 6. Description of projective shape spaces, even of finite configurations is a complicated task. Projective shape analysis, including distribution parametric and nonparametric approaches of projective shape can be performed in the context of multivariate axial data analysis [24]. For recent results in this area see also Lee et al. [20], Sughatadasa [34] and Liu et al. [22].

Acknowledgments

The authors wish to thank Texas Tech graduate student Jing Wang for providing the "Bigfoot" data set. The authors are also grateful for support from the National Science Foundation and the Deutsche Forschungsgemeinschaft (German Research Foundation). The authors are also very grateful to the two anonymous reviewers for their valuable comments which led to improvements of this paper.

References

[1] J.O. Berger, M. Delampady, Testing precise hypotheses, Statist. Sci. 2 (1987) 317–352.
[2] R.N. Bhattacharya, M. Denker, Asymptotic Statistics, DMV Seminar, vol. 14, Birkhäuser, Boston, 1990.
[8] J. Dauxios, A. Pousse, Y. Romain, Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference, J. Multivariate Anal. 12 (1982) 136–154.
[9] H. Dette, A. Munk, Validation of linear regression models, Ann. Statist. 26 (1998) 778–800.
[10] H. Dette, A. Munk, Some methodological aspects of validation of models in nonparametric regression, Statist. Neerlandica 57 (2003) 207–244.
[12] B. Efron, The Jackknife, the Bootstrap and Other Resampling Plans, CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 38, SIAM, Philadelphia, 1982.
[14] C. Goutis, C. Robert, Model choice in generalised linear models: a Bayesian approach via Kullback–Leibler projections, Biometrika 85 (1998) 29–37.
[16] J.L. Hodges, E.L. Lehmann, Testing the approximate validity of statistical hypotheses, J. Roy. Statist. Soc. B 16 (1954) 261–268.
[17] H. Hotelling, The generalization of Student's ratio, Ann. Math. Statist. 2 (1931) 360–364.
[19] E. Klassen, A. Srivastava, W. Mio, S. Joshi, Analysis of planar shapes using geodesic paths on shape spaces, IEEE T Pattern Anal. 26 (2004) 372–383.
[20] J.L. Lee, R. Paige, V. Patrangenaru, F. Ruymgaart, Nonparametric density estimation on homogeneous spaces in high level image analysis, in: R.G. Aykroyd, S. Barber, K.V. Mardia (Eds.), Bioinformatics, Images, and Wavelets, Department of Statistics, University of Leeds, 2004, pp. 37–40.
[21] J. Liu, B. Lindsay, Building and using semiparametric tolerance regions for parametric multinomial models, Technical Report, Penn State University, 2005.
[22] X. Liu, V. Patrangenaru, S. Sugathadasa, Projective shape analysis for noncalibrated pinhole camera views (To the Memory of W.P. Dayawansa), Technical Report M983, Florida State University, Department of Statistics (2007).
[23] Y. Ma, A. Soatto, J. Košecká, S. Sastry, An Invitation to 3-D Vision: From Images to Geometric Models, Springer, Berlin, 2005.
[24] K.V. Mardia, V. Patrangenaru, Directions and projective shapes, Ann. Statist. 33 (2005) 1666–1699.
[25] A. Munk, H. Dette, Nonparametric comparison of several regression functions: exact and asymptotic theory, Ann. Statist. 26 (1998) 2339–2368.
[26] R. Paige, V. Patrangenaru, F. Ruymgaart, W. Wang, Analysis of projective shapes of curves using projective frames, in: S. Barber, P.D. Baxter, K.V. Mardia, R.E. Walls (Eds.), Quantitative Biology, Shape Analysis, and Wavelets, Leeds University Press, Leeds, 2005, pp. 71–74.
[27] V. Patrangenaru, New large sample and bootstrap methods on shape spaces in high level analysis of natural images, Comm. Statist. Theory Methods 30 (2001) 1675–1695.

[28] J. Ramsay, B.W. Silverman, Functional Data Analysis, second ed., Springer, New York, 2005.

[29] S.N. Roy, R.C. Bose, Simultaneous confidence interval estimation, Ann. Math. Statist. 24 (1953) 513–536.

[30] F.H. Ruymgaart, S. Yang, Some applications of Watson's perturbation approach to random matrices, J. Multivariate Anal. 60 (1997) 48–60.

[31] A. Srivastava, S. Joshi, W. Mio, X. Liu, Statistical shape analysis: clustering, learning and testing, IEEE Trans. Pattern Anal. Mach. Intelligence 27 (4) (2005) 590–602.

[34] S.M. Sughatadasa, Affine and projective shape analysis with applications, Ph.D. Dissertation, Texas Tech University, 2006.

[35] G.S. Watson, Statistics on Spheres, Wiley, New York, 1983.